TRI DATA RELEASE PAPER
for Stakeholder Outreach

I. BACKGROUND

This is one of three papers which describe aspects of the Toxics Release Inventory (TRI) Program and raise issues for stakeholder input. The scope of each paper corresponds to a phase of the annual TRI reporting cycle. TRI data for a calendar year must be reported to EPA each year by July 1st after the end of the year. Therefore, reporting years are the same as calendar years. The "reporting cycle" begins with EPA's compliance assistance activities, including the development of its reporting forms and instructions package that is mailed to facilities in approximately March of each year. Once EPA receives the forms, it enters the data from the forms (over 91,000 in 2000) in its TRI database. After entry into the database, EPA runs a number of data quality checks on both the facility identification information and on the chemical-specific data. After the data entry and data quality steps are completed, the TRI database is "frozen" for analysis and development of data products for release to the public. Generally, EPA announces the annual release of the TRI data by holding a press event or issuing a press release, and simultaneously notifying a wide range of stakeholders.

The background paper for this stakeholder process is entitled *TRI Data Collection, Processing and Management,* and addresses the TRI data process beginning with submission of the forms and ending at the data "freeze." This paper, *TRI Data Release Issue Paper* is the second of the three; it discusses TRI data products, the process for analyzing and releasing the TRI data, uses of the data, and issues and considerations associated with these aspects of the TRI program. The third paper is *TRI Compliance Assistance Activities.* TRI compliance assistance activities are carried out throughout the year with certain of the activities being closely aligned with the reporting cycle.

As noted above, each year, once EPA finishes processing the TRI data it receives the previous July 1, the Agency "freezes" the data in order to prepare the data for the TRI public data release. Once the TRI data is "frozen," it takes the Agency approximately 3 months to conduct the analyses, prepare the data release reports, and prepare the data access tools to be ready to "release" the data to the public. Providing quality assurance of data and quality management of the public information and data tools is a key component at this time.

EPA makes the TRI data available to the public in several different ways. First, the data are provided to the public on-line through two EPA data access tools: TRI Explorer and the EPA Envirofacts on-line data warehouse. Second, the Agency prepares a two-volume Public Data Release (PDR) report and an Executive Summary which are made available in both hard copy and on the TRI website. Additional background materials, including charts and graphs, state and national data files,

press materials, etc., are also made available on the TRI website.

While some stakeholders have voiced concerns that EPA is not providing the data early enough, others have emphasized the need for the Agency to analyze the data and provide the type of context contained in the PDR reports at the same time EPA makes the TRI data publicly available each year. Undertaking these analyses and preparing the reports in a manner consistent with EPA quality systems affects the timing of the release of the data.

The purpose of this issue paper is to describe the data release process and how the TRI data is used, in order to solicit stakeholder input on the process and the data products. The next section of the paper, *TRI Data Products*, describes the many reports and data access tools EPA makes available for the TRI data release each year. The following section, entitled *TRI Data Release Process* describes how the data products are prepared, including a schedule of the tasks involved. The last section discusses several important aspects of the data release process (following the data freeze), such as timing of the release, the need for context for the data, the extent of data quality efforts, etc., as well as options for changing these aspects of the release. In each section we have included questions in italics to help focus the stakeholder comments.

II. TRI DATA PRODUCTS

Public Data Release Reports

Beginning in 1988, EPA has prepared a Public Data Release Report, or set of reports, each year in order to summarize the TRI data. The last 5 years' data release reports are available at www.epa.gov/tri/tridata/index.htm#pdr. The reports can also be obtained by calling the TRI Documents distribution line at (202) 564-9554.

In the last several years the Agency has been preparing three reports for the TRI data release. One is the Public Data Release Report, which provides a variety of analyses of the data including industry analyses, trend analyses, analyses of the data by media, by state, and by chemical. It also includes a large chemical table which summarizes the releases for each chemical for the current year and select previous years, including the first year of reporting on the chemical. The first chapter of this report explains the TRI reporting requirements and how they have changed over time. It also describes each data element and explains the factors to consider when using the data. These reports are roughly 400 pages long, including appendices that are approximately 200 pages. They contain over 150 tables and figures.

When the Agency is reporting on a set of chemicals or industries for the first time, the data release report will usually include more in-depth analyses of those industry sectors or chemicals. For

instance, this year's Public Data Release report on the 2000 TRI data included an extensive section on persistent bioaccumulative toxic chemicals, for which reporting requirements were revised in 2000.

The second major information product prepared each year is the "State Fact Sheets" report, which is a compilation of 4-page fact sheets on the TRI data for each state as well as summary tables which present state rankings on a variety of measures including total releases, on-site releases, off-site releases, total waste managed, releases from manufacturing, and from non-manufacturing industries. The individual state fact sheets include lists of top- ranked facilities for each state.

The third information product is an executive summary which is approximately 12 pages and provides an overview of the data, including summary tables, charts and graphs showing the distribution of releases and total wastes by industry, by media, and over time. It includes a short explanation of the TRI reporting requirements and the factors to be considered when using the data.

In the past few years, in efforts to make the TRI data available as early as possible, information products have been made available on the internet at the time the data is released and the printed reports have been made available approximately 6 weeks later. Thus, the Agency is able to release the TRI data 6 weeks earlier.

EPA would like comments on each of the data release reports, including how they meet users' data and information needs, and suggestions for changes.

TRI Electronic Data Access Tools

EPA has two main TRI electronic data access tools, the *TRI Explorer* (www.epa.gov/triexplorer/chemical.htm) and *Envirofacts* (www.epa.gov/enviro). Both can be accessed through the TRI web site at www.epa.gov/tri. *TRI Explorer* is a user-friendly tool that enables the user to generate reports on chemical releases and other waste management activities by facilities, chemicals, geographic areas, or industry type at the county, state, or national level. *Envirofacts* is EPA's on-line data "warehouse" which provides the public with direct access to EPA databases including TRI. Users can access basic facility information and chemical reports, which tabulate air emissions, surface water discharges, releases to land, underground injections, and transfers to off-site locations. *Envirofacts* also allows for custom queries, allowing the user to focus on specific data elements of interest.

In addition to these data access tools, EPA prepares TRI data files (in ASCII) that more sophisticated users can take and analyze themselves using data management and analytical tools of their choice, e.g., spreadsheets or other data management tools. These TRI data files are made available on the TRI web site (www.epa.gov/tri/tridata/state_data_files.htm). They are available by state or on a

national basis.

EPA is seeking users' comments on these data access tools, including comments on how they do or don't meet user needs. Please be as specific as possible about the data access tool(s) you are commenting on, and about any suggestions for improvements.

There are two other tools for accessing TRI data which have been developed by other organizations. One is *RTKNET* (www.rtk.net), developed by the organization OMB Watch. It provides access to a number of databases, including TRI, for communities to learn about their environment. The other is *TOXNET*, a National Library of Medicine tool to search the TRI data on-line (www.toxnet.nlm.nih.gov). Before EPA had its own electronic data access tools, *TOXNET* provided the only electronic access to the TRI data.

Other TRI Data Products

When the TRI data is made available to the public each year, the Agency creates a Data Release web page (www.epa.gov/tri/tridata/tri00/index.htm) that provides users with the TRI reports, access to *TRI Explorer* and *Envirofacts*, and a number of other information products that help summarize and provide context for the data. These include a set of press materials which contain summary charts, graphs, and tables, an overview of the TRI program which includes an explanation of the limitations of the data, Q's and A's, and other information that provides context to the TRI data.

EPA would like comments on these additional data tools in terms of how well they meet users' needs.

EPA is also seeking comments on the whole suite of tools made available for accessing and using the TRI data. Do these tools as a whole meet the wide range of user needs? Do users need more links to other data, including exposure, health, and environmental data, in order to effectively utilize the TRI data? Is more context information needed? Are there ways EPA can partner with other organizations to produce new tools?

III. TRI DATA RELEASE PROCESS

Development of all the data products that have been described here takes approximately 3 months, with many of the project tasks being done simultaneously. EPA has contractor support for preparing both the data release reports and the on-line data access tools. The most time-consuming part of the process of analyzing and releasing the data is the preparation of the two-volume data release reports, which take the entire three months, including review time. Figure 1 (attached) lays out the steps required to prepare the data release reports and the *TRI Explorer* data access tool. Many of the

steps are performed simultaneously because of the need to release the data as soon as possible.

One of the first steps in the process is preparing an outline and list of tables and figures to be included in the data release reports. This is done prior to the data freeze. Following the data freeze, the first major task involved with preparing the data release reports is to develop the data tables for preparing the public data release reports and for checking the *TRI Explorer*. These are necessary to analyze the data and write up the results. These tables are submitted to EPA for review. Preparing a first draft of the entire Public Data Release report takes about 8 weeks from the time the data is frozen (some of the text is prepared before all the tables are run). After review and comment by EPA the draft needs to be sent to the design subcontractor for layout. This step takes about 1 week. Finally approximately 2 weeks are needed for review and comment by the Agency.

The State Fact Sheets and the ranking tables contained in the State Fact Sheets Report take roughly 2 weeks to prepare in draft form, and then a week is required for formatting and layout by the design subcontractor. Then they are submitted to EPA for review both by EPA and the states. EPA has the regional TRI coordinators distribute individual state fact sheets to the TRI coordinators in their region to review and make sure they are consistent with their databases. This review/correction process takes approximately 2.5 weeks. Therefore, a total of about 5-6 weeks is necessary for preparation of the State Fact Sheets Report.

The Executive Summary takes approximately 6 weeks to prepare, including approximately 2 weeks for analyzing the data and developing the tables and 2 weeks for the text. In addition, approximately 2 weeks is needed for review and design work on the report. Finally a week is needed for printing.

Timeframes for the development of the additional data release materials are provided in Figure 1, i.e. roughly 1.5 weeks for the contractor to develop press materials, 4 weeks for EPA to create the national and state data files in ASCII, 6 weeks for EPA to develop and coordinate Agency review of a set of Qs and As on the data (once the tables are prepared), and 5 weeks to prepare materials for notifying interested parties of the data release.

Preparation of the *TRI Explorer* for the TRI public data release varies depending upon whether enhancements are being introduced at the time of the data release. If no new enhancements are being introduced, it should take approximately 2 months to prepare the *TRI Explorer* for the data release. If new enhancements are being introduced, additional time is needed for developing, testing, and correcting the tool, increasing the total time needed to about 2.5 - 3 months.

IV. OPTIONS FOR RELEASING THE TRI DATA EARLIER

There are ways in which the Agency could reduce the amount of time it takes to make the data available after the data freeze. One way would be to simply make the data available through *Envirofacts* and *TRI Explorer*. This would take approximately 6-7 weeks from the data freeze. However, it would not allow the Agency time to analyze the data in order to answer questions that might be raised nor time to prepare contextual materials other than the standard documents that explain the data and its limitations. In addition, it would eliminate part of the data quality efforts that are carried out during this time period.

Another option for reducing the amount of time it takes to release the data would be to prepare the Executive Summary, as is currently done, along with the release of the data on *TRI Explorer* and *Envirofacts*. This would take approximately 8-9 weeks from the data freeze and would allow a little time to review the data and to begin to anticipate questions. However, very little analysis could be done in this time. In both cases, the complete Public Data Release Reports could be completed and released after the release of the data.

V. ISSUES AND CONSIDERATIONS

Timing

All things being equal, we would clearly like to make the TRI data available to the public as early as possible since data is more useful the more current it is. Since the reporting deadline is July 1 following the year for which the data is reported, there is a statutory 6 month delay before the data are submitted to EPA and the states. As is discussed in more detail in the paper describing the processing of the TRI data, the Agency takes a few months to enter the data and undertake quality assurance measures before freezing the data. This paper has discussed the process the Agency traditionally goes through to analyze the data and prepare reports summarizing and explaining the data. In order to release the data sooner after freezing the data, we would need to forego some or all of the analysis and report-writing, which could have implications for the quality of the data, and the amount of analysis and context made available to the public.

EPA is seeking comment on the timing of the data release, and on the preferred tradeoffs that should be made in order to change the timing of the data release. For instance, if a commenter feels strongly that the TRI data needs to be released at least 3 months earlier, we would like the commenter's suggestions regarding how to accomplish that, e.g., specific recommendations on scaling back or eliminating the data release reports

Context

EPA has always prepared its data analysis and some type of report prior to releasing the data.

In addition, some type of press announcement is always made at the time of the release to make people aware that the data is available. Over the years, many stakeholders have indicated that they feel it is important for the Agency to provide context when releasing the data since, otherwise, there could be a lot of room for misinterpretation. Providing this context adds to the time it takes to make the data available.

EPA is seeking comment on the context currently provided when we release the TRI data, including specific recommendations for change. For instance, if you would like to see more context provided, please provide recommended additions.

Extent of Data Analysis

Given the huge volume of data in the TRI database, there is much analysis that can be done on it. However, the Agency is limited in the amount of analysis that can be done on the data because of the need to make the data available as quickly as possible. Therefore, there are many questions about the data which cannot be fully answered at the time of the data release. If we consider trying to accelerate the data release schedule, we need to recognize the fact that the Agency's analysis of the data will be even more limited than it has been. Therefore, questions may be raised about the data that cannot be answered immediately.

EPA is seeking comment on the extent of the data analysis currently conducted as well as specific recommendations for change. If you are recommending that the Agency conduct additional analysis prior to releasing the data, please comment on your willingness to have the data released later than it currently is.

Data Quality

Assuring the quality of the TRI data has been, and continues to be an important aspect of the TRI Program. The utility of the data is dependent upon how credible it is. The most important data quality activities are discussed in the other issue papers. For example, the compliance assistance issue paper discusses the many TRI guidance documents, training workshops, the hotline, etc. The other issue paper addressing TRI data receipt and processing discusses several data quality steps taken prior to the data freeze. In addition, as the data is analyzed following the data freeze, additional quality assurance is undertaken. At this stage in the process, it is very difficult for the Agency to change the data since it has already been "frozen" and generally many analyses have already been performed. Therefore, if significant errors are discovered, the remedy is usually to explain them in a footnote.

If EPA were to try and speed the release of the TRI data following the data release, data quality could be compromised to some extent. For example, in the process of developing the Public

Data Release and State Fact Sheets reports, EPA occasionally discovers significant errors, mostly resulting from erroneous submissions. If EPA were to simply make the data available through Envirofacts and TRI Explorer, it's unlikely these types of errors would be caught, partly due to the fact that we would not be having the states check the State Fact Sheet data.

Options for expediting the TRI data release are likely to include eliminating or reengineering some of the data quality activities, and placing greater responsibility on the reporting facilities to submit their data correctly in the first place. This could compromise data quality in the short run as we move to the new approach, but could benefit the program in the long run.

EPA is seeking comment on users' trade-off preferences between data quality and timing of the data release. For instance, is expediting the release of the data important enough to users that they are willing to tolerate the possibility that there will be more errors in the data at the time of release?

Digital Divide

Given budget constraints, the growth of the Internet, and the pressure to make information available as quickly as possible, we have become more and more reliant on the Internet to make the TRI data available. We need to keep in mind that there are still many citizens that need the TRI data who cannot access it electronically very easily. Many of the options for expediting the release of the data rely exclusively on electronic data access.

EPA is seeking comment on whether or not the digital divide is a significant issue, and if it is, how EPA should be addressing it with respect to the TRI public data release.